A Riemannian Optimization Technique for Rank Inequality Constraints

Guifang Zhou[1], Wen Huang[2], K. A. Gallivan[1], Paul Van Dooren[2], P.-A. Absil[2]
[1]Florida State University, [2]Université Catholique de Louvain

This research was supported by the National Science Foundation under grant NSF-1262476. This paper presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme initiated by the Belgian Science Policy Office. This work was supported by grant FNRS PDR T.0173.13.

Problem and Applications

This study considers combining rank inequality constraints with a matrix manifold constraint in a problem of the form

$$\min_{x \in \mathcal{M}_{\leq k}} f(x), \tag{1}$$

where $\mathcal{M}_{\leq k} = \{x \in \mathcal{M} | \mathrm{rank}(x) \leq k\}$ and $\mathcal{M}$ is a submanifold of $\mathbb{R}^{m\times n}$. Numerous applications exist, e.g., [ZW03, FHB04, MLP+06, JHSX11].
The details of this work can be found in [ZHG+15].

Background

Riemannian optimization methods play important roles:

- $\mathcal{M} = \mathbb{R}^{m\times n}$ in most of applications;
- $\mathbb{R}_r^{m\times n} := \{x \in \mathbb{R}^{m\times n} | \mathrm{rank}(x) = r\}$ is a Riemannian manifold.

Existing methods choose the k in (1) a priori. However, it is not easy to choose a suitable k.

- The solution with too small k may be unacceptable;
- The computational time may be unacceptable with too large k.

Contribution

- Generalize the admissible set from $\mathbb{R}^{m\times n}_{\leq k}$ to $\mathcal{M}_{\leq k}$;
- Define an algorithm solving a rank inequality constrained problem while finding a suitable rank for approximation;
- Prove theoretical convergence results;
- Implementations based on Riemannian optimization methods.

Basic Idea

Apply Riemannian optimization methods on a fixed rank manifold $\mathcal{M}_r$ while efficiently and effectively updating the rank r.

Increase Rank

Increase rank if next two conditions hold.

- Condition I (angle threshold θ_0):

$$\angle(\mathrm{grad} f_{\mathrm{F}}(x_r), \mathrm{grad} f_r(x_r)) = \theta > \theta_0,$$

- Condition II (difference threshold, ϵ_2):

$$\|\mathrm{grad} f_{\mathrm{F}}(x_r) - \mathrm{grad} f_r(x_r)\| \geq \epsilon_2,$$

where $x_r \in \mathcal{M}_r$, $\mathrm{grad} f_{\mathrm{F}}(x)$ and $\mathrm{grad} f_r(x)$ are the Riemannian gradients with respect to $\mathcal{M}$ and $\mathcal{M}_r$ respectively.

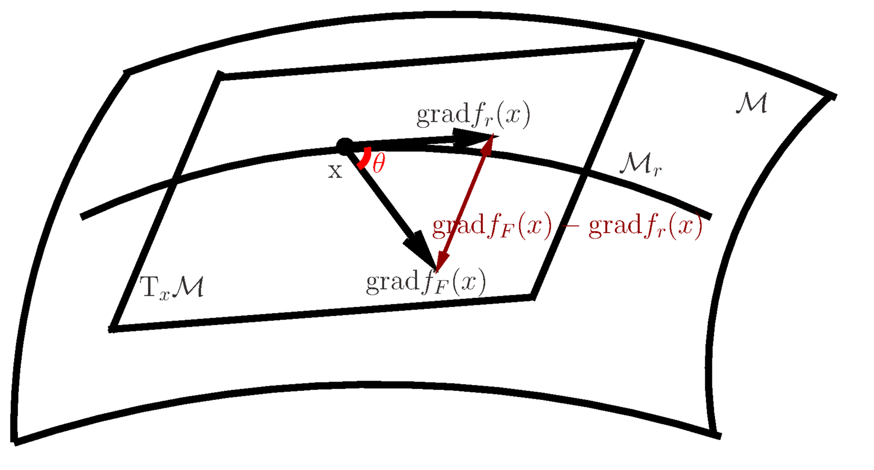

Figure 1. Strategy of increasing the rank.

Rank-Related Objects

The new concepts of rank-related vector and rank-related retraction play an important role in updating the rank and avoiding increasing it excessively.

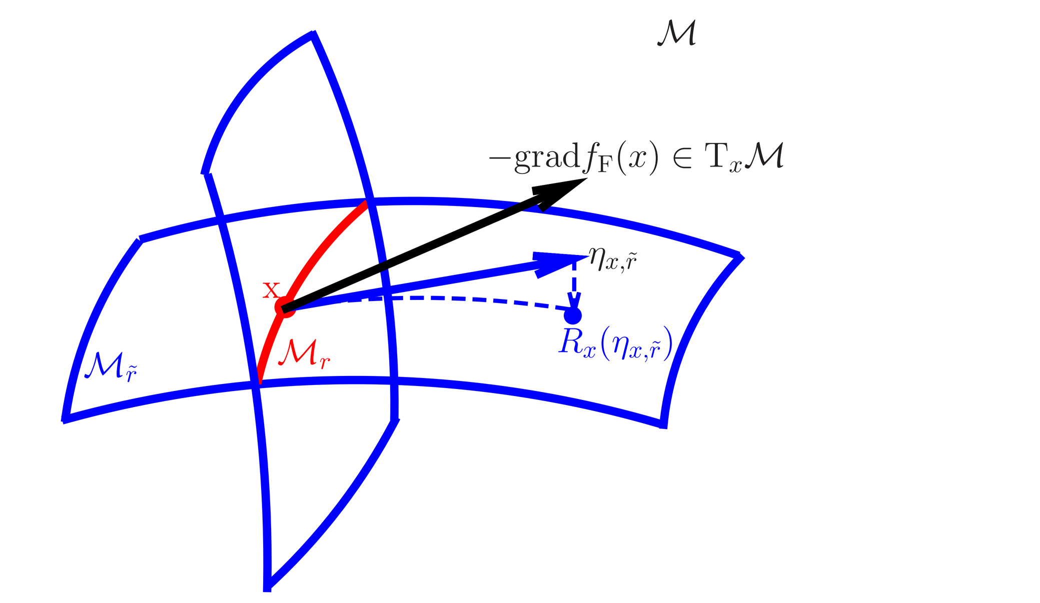

Figure 2. $x \in \mathcal{M}_r$; $r < \tilde{r}$; $\eta_{x,\tilde{r}}$ is a rank-$\tilde{r}$-related vector, i.e., there exists a curve $\gamma(t)$ such that $\gamma(0) = x$, $\dot{\gamma}(0) = \eta_{x,\tilde{r}}$ and $\mathrm{rank}(\gamma(t)) = \tilde{r}$; R is a rank-related retraction, i.e., $\mathrm{rank}(R_x(t\eta_{x,\tilde{r}})) = \tilde{r}$ for $t \in (0, \delta)$, $\delta > 0$.

Reduce Rank

- Truncate rank and retract to $\mathcal{M}$
- May increase the function value
- Do not destroy the progress that is made in the previous rank increasing step, i.e.,

$$f(X_{i+1}) - f(X_s) \leq c(f(X_{s+1}) - f(X_s)),$$

where s is such that the latest rank increase was from X_s to X_{s+1}, $0 < c < 1$.

Algorithm

Algorithm 1

for n = 0,1, 2,. . . **do**
  Approximately optimize f over $\mathcal{M}_r$ with initial point x_n and obtain $\tilde{x}_n$;
  if $\tilde{x}_n$ is not close to $\mathcal{M}_{\leq r-1}$ **then**
    if Both Conditions I and II are satisfied **then**
      Find the updated rank $\tilde{r} \in (r, k]$ such that $\|\,\mathrm{grad}\, f_F(\tilde{x}_n) - \eta^*\|_F$ is sufficiently small, where $\eta^* \in \arg\min_{\eta \in \mathrm{T}_{\tilde{x}_n}\mathcal{M}_{\leq \tilde{r}}} \|\,\mathrm{grad}\, f_F(\tilde{x}_n) - \eta\|_F$ is a rank-$\tilde{r}$-related vector.
      Obtain x_{n+1} by applying a line search algorithm along η^* using a rank-related retraction;
    else
      If x_{n+1} is accurate enough, stop.
    end if
  else
    Reduce the rank of $\tilde{x}_n$ if the progress made in the previous rank increasing step is not destroyed; Update r; Obtain next iterate x_{n+1};
  end if
end for

Main Theoretical Results

Under some reasonable assumptions:

- (Global Result) The sequence $\{x_n\}$ generated by Algorithm 1 satisfies $\liminf_{n\to\infty} \|P_{\mathrm{T}_{x_n}\mathcal{M}_{\leq k}}(\mathrm{grad} f_{\mathrm{F}}(x_n))\| \leq \left(\sqrt{1+\frac{1}{\epsilon_1^2}}\right)\epsilon_2$, where $\epsilon_1 = \tan(\theta_0)$.
- (Local Result) The sequence $\{x_n\}$ enters a neighborhood $\mathcal{U}_*$ of a minimizer x_* and remains in $\mathcal{U}_*$. The distance $\mathrm{dist}(x_n, x_*)$ is bounded based on ϵ_1, ϵ_2 and $\mathrm{Hess}\, f_{\mathrm{F}}(x_*)$. The ranks of $\{x_n\}$ are fixed eventually.

Weighted Low Rank Problems

Weighted low rank problem concerns solving

$$\min_{X \in \mathcal{M}_{\leq k}} \|A - X\|_W^2$$

where $\mathcal{M} = \mathbb{R}^{m\times n}$, A is given, $W \in \mathbb{R}^{mn\times mn}$ is symmetric positive definite and $\|A - X\|_W^2 = \mathrm{vec}(A - X)^T W \mathrm{vec}(A - X)$.

Experiments

Algorithm 1 is compared with the state-of-the-art methods for weighted low rank approximation problems.

The matrix A is generated by $A_1 A_2^T \in \mathbb{R}^{80\times 10}$, where $A_1 \in \mathbb{R}^{80\times 5}$, $A_2 \in \mathbb{R}^{10\times 5}$. $W = U\Sigma U^T$ where $U \in \mathbb{R}^{800\times 800}$ is a random orthogonal matrix generated by Matlab's RAND and QR. The 800 singular values of W are generated by Matlab LOGSPACE with condition number 100 and mutiplying, element-wise, by a uniform distribution matrix on the interval [0.5, 1.5].

	k	rank	f	R_err	time(s)
(1)	3	3	8.65_{1}	3.10_{-1}	6.16_{-1}
	5	5	3.29_{-22}	1.09_{-13}	5.19_{-1}
	7	5	3.65_{-17}	1.87_{-10}	4.43_{-1}
(2)	3	3	8.65_{1}	3.10_{-1}	3.03
	5	5	1.25_{-20}	3.45_{-12}	3.05
	7	$5.02_{99\%}$	2.13_{-17}	3.740_{-11}	1.63
(3)	3	3	8.65_{1}	3.10_{-1}	3.96
	5	5	3.05_{-12}	5.41_{-8}	1.06
	7	$7_{0\%}$	2.23_{-12}	4.77_{-8}	2.07
(4)	3	3	8.65_{1}	3.10_{-1}	9.11
	5	5	8.28_{-10}	8.93_{-7}	4.05
	7	$7_{0\%}$	2.34_{-10}	4.86_{-7}	6.89

Table 1. An average of 100 random runs. (1), (2), (3) and (4) denote Alg. 1, DMM [BM06], SULS [SU14] and APM [LPW97] respectively. ϵ_1 and ϵ_2 are chosen to be $\sqrt{3}$ and 10^{-4} for Algorithm 1. R_err denotes $\|A - X\|_W / \|A\|_W$. The subscript $\pm k$ indicates a scale of $10^{\pm k}$. The subscript $m\%$ denotes that the percentage of runs that find the true rank.

references

[BM06] I. Brace and J. H. Manton. An improved BFGS-on-manifold algorithm for computing weighted low rank approximations. *Proceedings of 17th international Symposium on Mathematical Theory of Networks and Systems*, pages 1735–1738, 2006.
[FHB04] M. Fazel, H. Hindi, and S. Boyd. Rank minimization and applications in system theory. *Proceedings of American Control Conference*, 2004.
[JHSX11] H. Ji, S.-B. Huang, Z. Shen, and Y. Xu. Robust video restoration by joint sparse and low rank matrix approximation. *SIAM Journal on Imaging Sciences*, pages 1–19, 2011.
[LPW97] W.-S. Lu, S.-C. Pei, and P.-H. Wang. Weighted low-rank approximation of general complex matrices and its application in the design of 2-d digital filters. *IEEE Transactions on Circuits and Systems*, 44(7):650–655, 1997.
[MLP+06] I. Markovsky, M. Luisa Rastello, A. Premoli, A. Kukush, and S. Van Huffel. The element-wise weighted total least-squares problem. *Computational Statistics & Data Analysis*, 50(1):181–209, January 2006. doi:10.1016/j.csda.2004.07.014.
[SU14] R. Schneider and A. Uschmajew. Convergence results for projected line-search methods on varieties of low-rank matrices via lojasiewicz inequality. February 2014. 1402.5284.
[ZHG+15] G. Zhou, W. Huang, K. A. Gallivan, P. Van Dooren, and P.-A. Absil. Rank-constrained optimization: A Riemannian manifold approach. In *In Proceeding of 23th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, number 1, pages 249–254, 2015.
[ZW03] Z. Zhang and L. Wu. Optimal low-rank approximation to a correlation matrix. *Linear Algebra and its Applications*, 364:161–187, May 2003. doi:10.1016/S0024-3795(02)00551-7.